REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
24-06-2008	Final Report	15-Dec-2005 - 14-Jun-2008

4. TITLE AND SUBTITLE
Final Report of "Next-Generation Botnet Detection and Response"

5a. CONTRACT NUMBER
W911NF-06-1-0042

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
5M30V2

6. AUTHORS
Wenke Lee

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Georgia Institute of Technology
Office Of Contract Administration
Program Initiation Division
Atlanta, GA 30332 -0420

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
49629-CI-DRP.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
In this project, we developed dynamic DNS monitoring heuristics to identify domains used for botnet command and control, as well as anomaly detection algorithms for Recursive DNS servers at ISPs and enterprise networks to detect botnet command and control activities. We also developed botnet detection systems for enterprise networks. These systems include BotHunter, BotSniffer, BotMiner, and BotProbe.

We formed a start-up company Damballa, Inc. to deliver anti-botnet technologies to government and enterprise customers.

15. SUBJECT TERMS
botnet detection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Wenke Lee
U	U	U	SAR		19b. TELEPHONE NUMBER 404-385-2879

Report Title

Final Report of "Next-Generation Botnet Detection and Response"

ABSTRACT

In this project, we developed dynamic DNS monitoring heuristics to identify domains used for botnet command and control, as well as anomaly detection algorithms for Recursive DNS servers at ISPs and enterprise networks to detect botnet command and control activities. We also developed botnet detection systems for enterprise networks. These systems include BotHunter, BotSniffer, BotMiner, and BotProbe.

We formed a start-up company Damballa, Inc. to deliver anti-botnet technologies to government and enterprise customers.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Number of Papers published in peer-reviewed journals: 0.00

(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none)

Number of Papers published in non peer-reviewed journals: 0.00

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

Peer-Reviewed Conference Proceeding publications (other than abstracts):

1. Modeling Botnet Propagation Using Time Zones. ?David Dagon, Cliff Zou, and Wenke Lee. ?In Proceedings of The 13th Annual Network and Distributed System Security Symposium (NDSS 2006), San Diego, CA, February 2006.

2. BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation. ?Guofei Gu, Phillip Porras, Vinod Yegneswaran, Martin Fong, and Wenke Lee. ?In Proceedings of The 16th USENIX Security Symposium (Security'07), Boston, MA, August 2007.

3. A Taxonomy of Botnet Structures. ?David Dagon, Guofei Gu, Chris Lee, and Wenke Lee. ?In Proceedings of The 23rd Annual Computer Security Applications Conference (ACSAC 2007), Miami Beach, FL, December 2007.

4. Corrupted DNS Resolution Paths: The Rise of a Malicious Resolution Authority. David Dagon, Chris Lee, Niels Provos, and Wenke Lee. In Proceedings of The 15th Annual
Network and Distributed System Security Symposium (NDSS 2008), San Diego, CA, February 2008.

5. BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic. Guofei Gu, Junjie Zhang, and Wenke Lee. In Proceedings of The 15th Annual Network and Distributed System Security Symposium (NDSS 2008), San Diego, CA, February 2008.

6. BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection. Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee. In Proceedings of The 17th USENIX Security Symposium (Security'08), San Jose, CA, August 2008.

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts): 6

(d) Manuscripts

Number of Manuscripts: 0.00

Number of Inventions:

Graduate Students

NAME	PERCENT_SUPPORTED
David Dagon	0.50
Junjie Zhang	0.50
Kapil Singh	0.50
FTE Equivalent:	**1.50**
Total Number:	**3**

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Roberto Perdisci	0.50
FTE Equivalent:	**0.50**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Wenke Lee	0.25	No
FTE Equivalent:	**0.25**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
Robert Edmonds	0.25
FTE Equivalent:	**0.25**
Total Number:	**1**

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	1.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	1.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	1.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME PERCENT_SUPPORTED

FTE Equivalent:

Total Number:

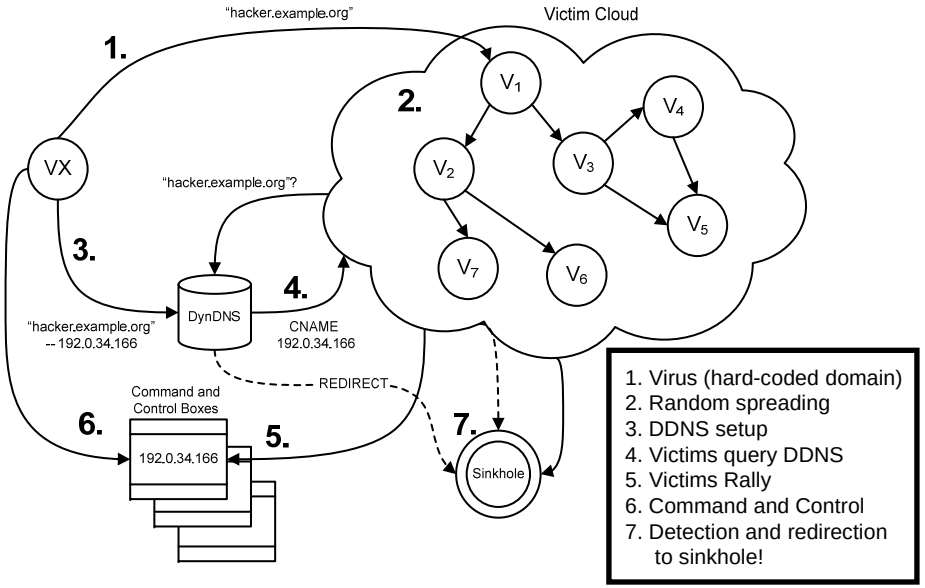
Sub Contractors (DD882)

Inventions (DD882)

Next-Generation Botnet Detection and Response

Georgia Institute of Technology

Start-Date, Dec. 2005 **Email:** wenke@cc.gatech.edu **WWW:** http://www.cc.gatech.edu/~wenke Dec. 2005



Project Objectives

• Develop technologies to identify and remediate attacking networks (e.g., botnets).

• Disrupt the botnet command and control (C&C). Without C&C, a botnet is an unorganized infection.

• Detection techniques must be evasion-resistant and not dependant on one given protocol.

Scientific/Technical Approaches

• *DNS-Based Detection*: Using DDNS and high-speed DNS monitoring, we will detect botnet activity, regardless of the underlying C&C protocol.

• *Flow/traffic-Based Detection:* We will use flow-based anomaly detection techniques for evasive botnets that don't even use DNS.

• *Response*: We will use proxynets, blackholes, sinkholes and other technologies to disrupt the botnet C&C, and enable traditional response techniques.

Accomplishments

• Developed and deployed a set of DNS based monitoring and surveying systems for Internet-scale botnet detection and situation awareness.

• Developed a family of botnet detection systems for enterprise networks.

• On-going and successful technology transfer: Damballa.

• New project from DHS: prototype and deployment.

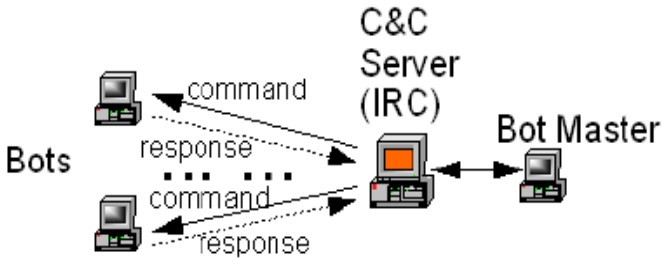
Next-Generation Botnet Detection and Response

- Highlights
  - Dynamic DNS monitoring heuristics to identify domains used for botnet command and control
  - Surveying method for (misconfigured/malicious) Open Recursive DNS servers on the Internet
  - Anomaly detection algorithms for Recursive DNS servers at ISPs and enterprise networks
  - Botnet detection systems for enterprise networks
    - BotHunter, BotSniffer, BotMiner, and BotProbe
  - Related efforts
    - CyberTA (SRI), new DHS project
  - Formed a start-up company Damballa, Inc. to deliver anti-botnet technologies to government and enterprise customers.

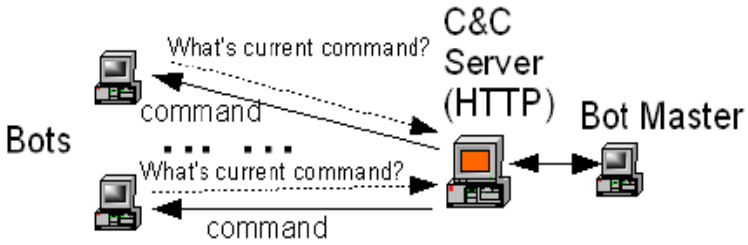
We highlight the BotSniffer system in this report. We provide a list of publications at the end of this report. These papers describe the technologies developed in this project in great details.

BotSniffer: Detecting Botnet C&C in Enterprise Netowrks

Botnet C&C Communication

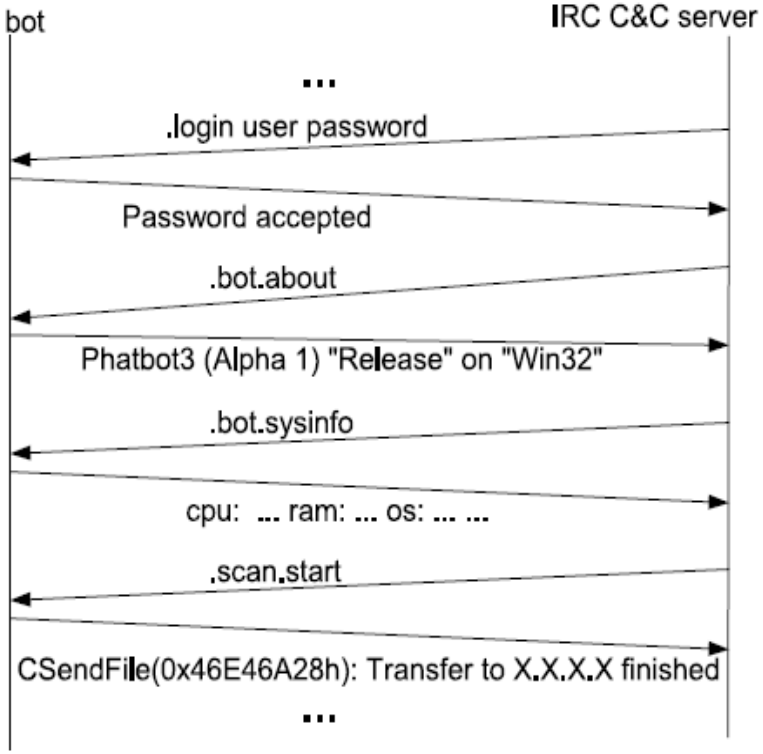


(I) C&C: Push style



(II) C&C: Pull style

(a) Two styles of botnet C&C

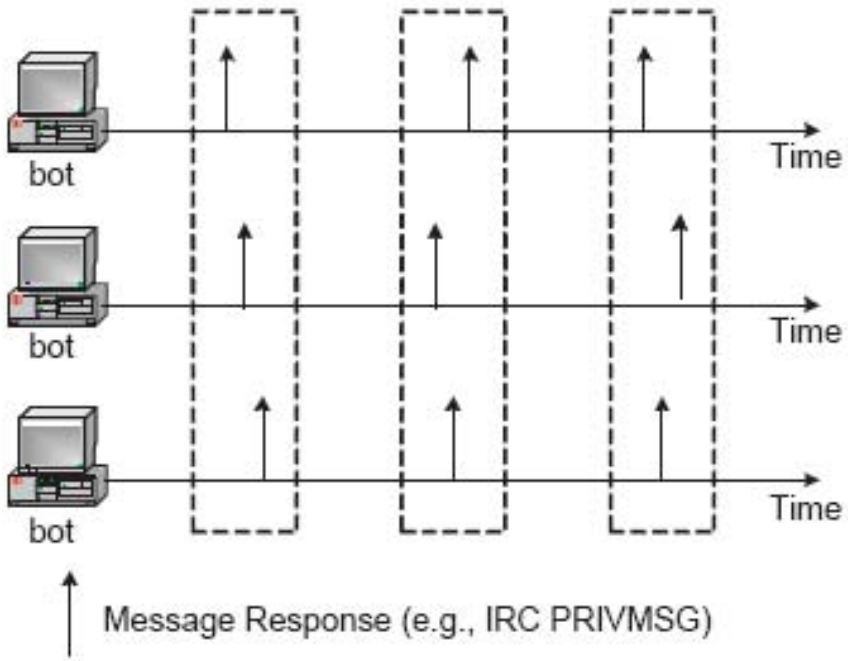


(b) An IRC-based C&C communication example

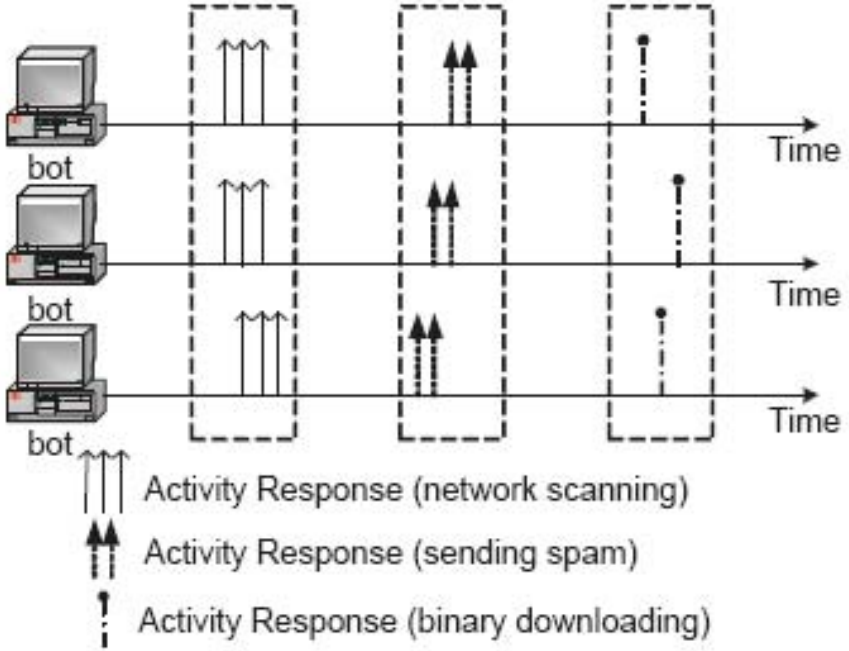
Botnet C&C Detection

- C&C is essential to a botnet
  - Without C&C, bots are just discrete, unorganized infections
- C&C detection is important
  - Relatively stable and unlikely to change within botnets
  - Reveal C&C server and local victims
  - The weakest link if C&C server is detected and can be taken down
- C&C detection is hard
  - Use existing common protocol instead of new one
  - Low traffic rate
  - Obscure/obfuscated communication

Botnet C&C: Spatial-Temporal Correlation and Similarity

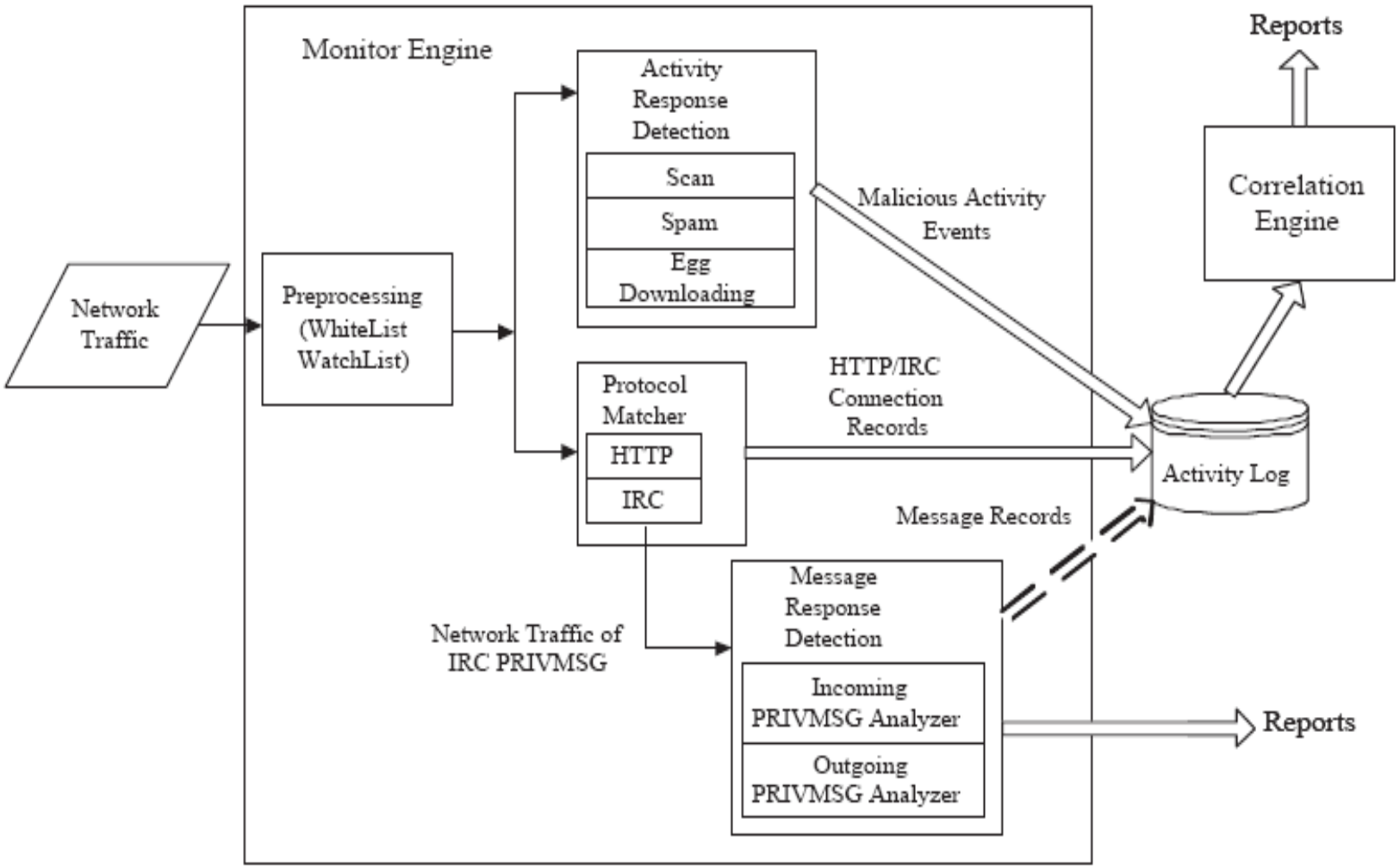


(a) Message response crowd



(b) Activity response crowd

BotSniffer Architecture



Correlation Engine

- Group clients according to their destination IP and Port pair (HTTP/IRC connection record)
- Perform a *group analysis* on spatial-temporal correlation and similarity property

- Currently
  - Response-Crowd-Density-Check algorithm for group activity response analysis
  - Response-Crowd-Homogeneity-Check algorithm for group message response analysis.

Response-Crowd-Density-Check Algorithm

- Response crowd
  - a set of clients having (message/activity) response behavior
- Dense crowd
  - the fraction of the number of such message/activity response clients in the crowd over the size of the group is larger than a threshold (e.g., 0.5)
- Example: 5 clients connected to the same IRC/HTTP server, and all of them scan at similar time (or send messages at similar time)
- Sequential Probability Ratio Testing

Sequential Probability Ratio Testing (SPRT)

- Each round (a time window), observe whether current crowd is dense or not (Y)
  - Hypothesis
    - Pr(Y=1|H1) very high (for botnet)
    - Pr(Y=1|H0) very low (for normal user)
- Make a random walk according to the observation Y
- After several rounds, we may reach a decision (which hypothesis is more likely, H1 or H0)

- Also called TRW (Threshold Random Walk)
- Bounded false positive and false negative rate (as desired), and usually needs only a few rounds

$$\Lambda_n = \ln \frac{Pr(Y_1, \ldots, Y_n | H_1)}{Pr(Y_1, \ldots, Y_n | H_0)} = \ln \frac{\prod_i Pr(Y_i | H_1)}{\prod_i Pr(Y_i | H_0)} = \sum_i \ln \frac{Pr(Y_i | H_1)}{Pr(Y_i | H_0)}$$

Response-Crowd-Homogeneity-Check Algorithm

- A homogeneous response crowd
  - most of the members have very similar responses
- Similarity is defined
  - Message response
    - Similar payload (Dice distance)

$$Dice(X,Y) = \frac{2|ngrams(X) \cap ngrams(Y)|}{|ngrams(X)| + |ngrams(Y)|}$$

  - Activity response
    - Scan same ports (subnet)
    - Download same binary
    - Send similar spam

Experiments

Trace	trace size	duration	Pkt	TCP flow	server	FP
IRC-1	54MB	171h	189,421	10,530	2,957	0
IRC-2	14MB	433h	33,320	4,061	335	0
IRC-3	516MB	1,626h	2,073,587	4,577	563	5
IRC-4	620MB	673h	4,071,707	24,837	228	2
IRC-5	3MB	30h	19,190	24	17	0
IRC-6	155MB	168h	1,033,318	6,981	85	1
IRC-7	60MB	429h	393,185	717	209	0
All-1	4.2GB	10m	4,706,803	14,475	1,625	0
All-2	6.2GB	10m	6,769,915	28,359	1,576	0
All-3	7.6GB	1h	16,523,826	331,706	1,717	0
All-4	15GB	1.4h	21,312,841	110,852	2,140	0

Experiments (cont.)

BotTrace	trace size	duration	Pkt	TCP flow	Detected
B-IRC-G	950k	8h	4,447	189	Yes
B-IRC-J-1	-	-	143,431	-	Yes
B-IRC-J-2	-	-	262,878	-	Yes
V-Rbot	26MB	1,267s	347,153	103,425	Yes
V-Spybot	15MB	1,931s	180,822	147,921	Yes
V-Sdbot	66KB	533s	474	14	Yes
B-HTTP-I	6MB	3.6h	65,695	237	Yes
B-HTTP-II	37MB	19h	395,990	790	Yes

Discussion & Future Work

- Evading HTTP autocorrelation by using very long period
- Evasion using other protocols or self-designed protocols
- Effect of encryption
- Evasion by using random delay/period, injecting random noise, injecting random garbage in the packet
- *A new system under development will address these problems*

Project Statistics and Summary

Students supported:

- 1 undergraduate student
- 3 graduate students
- 2 PhDs expected May/August 2008

Publications:

- 5 Conference papers
- 1 book chapter

Technology Transitions:

- 4 Patents (disclosures)
- 1 start-up: Damballa, Inc.
- 1 DHS Type II project

Publication list

1. Modeling Botnet Propagation Using Time Zones. David Dagon, Cliff Zou, and Wenke Lee. In *Proceedings of* ***The 13th Annual Network and Distributed System Security Symposium (NDSS 2006)***, San Diego, CA, February 2006.
2. BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation. Guofei Gu, Phillip Porras, Vinod Yegneswaran, Martin Fong, and Wenke Lee. In *Proceedings of* ***The 16th USENIX Security Symposium (Security'07)***, Boston, MA, August 2007.
3. A Taxonomy of Botnet Structures. David Dagon, Guofei Gu, Chris Lee, and Wenke Lee. In *Proceedings of* ***The 23rd Annual Computer Security Applications Conference (ACSAC 2007)***, Miami Beach, FL, December 2007.
4. Corrupted DNS Resolution Paths: The Rise of a Malicious Resolution Authority. David Dagon, Chris Lee, Niels Provos, and Wenke Lee. In *Proceedings of* ***The 15th Annual Network and Distributed System Security Symposium (NDSS 2008)***, San Diego, CA, February 2008.
5. BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic. Guofei Gu, Junjie Zhang, and Wenke Lee. In *Proceedings of* ***The 15th Annual Network and Distributed System Security Symposium (NDSS 2008)***, San Diego, CA, February 2008.
6. BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection. Guofei Gu, Roberto Perdisci, Junjie Zhang, and Wenke Lee. In *Proceedings of* ***The 17th USENIX Security Symposium (Security'08)***, San Jose, CA, August 2008.